International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 10-No.4 (2024) pp. 3960-3968
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Governance of Clinical Data Quality in Electronic Health Records: Administrative Policies, Health Informatics Systems, and Medical Secretarial Practices

Bader Abdulrahman Sulayman Almuballi[1], Muslih Quaymil Nughaymish Alshamlani[2]*, Yasir Abdullah Alghofaili[3], Faihan Hamdan Lattif Alshammari[4], Latifeh Matar Salem Alattwei[5], Abdullah Abdulrahman Alsaif[6], Amjad Mubarak Alansari[7], Orjwan Mansoor Alhazmi[8], Nouf Khalid Abualfaraj[9], Salma Omar Mandour[10]

[1]Health Informatics Technician, Maternity and Children Hospital in Hail, Hail Health Cluster, Ministry of Health, Hail Region, Saudi Arabia
Email: bade2r@gmail.com - **ORCID:** 0000-0002-8793-3862

[2]SMedical Records Technician, Prince Abdullah bin Abdulaziz bin Musaed Center for Cardiology and Cardiac Surgery, Arar, Northern Borders Health Cluster, Ministry of Health, Northern Borders Region, Saudi Arabia
* **Corresponding Author Email:** emuslihqa@moh.gov.sa - **ORCID:** 0000-0002-2567-7937

[3]Medical Secretary Technician, Prince Sultan Cardiac Center, Qassim Health Cluster, Ministry of Health, Buraydah, Qassim Region, Saudi Arabia
Email: yalghofaili@moh.gov.sa - **ORCID:** 0000-0002-2567-7937

[4]Health Informatics Technician, Maternity and Children Hospital in Hail, Hail Health Cluster, Ministry of Health, Hail Region, Saudi Arabia
Email: k5o5@hotmail.com - **ORCID:** 0000-0002-2567-7937

[5]Health Informatics Technician, Maternity and Children's Hospital, Tabuk, Tabuk Health Cluster, Ministry of Health, Tabuk Region, Saudi Arabia
Email: Lalattwei@moh.gov.sa - **ORCID:** 0000-0002-2567-7937

[6]Health Administration Specialist, Eradah Mental Health Complex, Hail, Hail Health Cluster, Ministry of Health, Hail Region, Saudi Arabia
Email: ababalsaif@moh.gov.sa - **ORCID:** 0000-0002-2567-7937

[7]Health Services and Hospital Management Specialist – Security Forces Hospital Makkah – Ministry of Interior – Makkah – Makkah Region – Saudi Arabia
Email: amalansari@sfhm.med.sa - **ORCID:** 0000-0002-2567-7937

[8]health informatics specialist- security force hospital makkah - ministry of interior- makkah region - Saudi Arabia
Email: omalhazmi@sfhm.med.sa- **ORCID:** 0000-0002-2567-7937

[9]health informatics specialist- security force hospital makkah - ministry of interior- makkah region - Saudi Arabia
Email: nabu-alfaraj@sfhm.med.sa - **ORCID:** 0000-0002-2567-7937

[10]Health Services and Hospital Management Specialist, Security Forces Hospital Makkah, Ministry of Interior, Makkah Region, Saudi Arabia
Email: Somandour@sfhm.med.sa - **ORCID:** 0000-0002-2567-7937

Article Info:

DOI: 10.22399/ijcesen.4768
Received : 01 October 2024
Accepted : 30 October 2024

Keywords

Clinical Data Quality,
Electronic Health Records (EHR),
Data Governance,
Health Informatics,
Health Information Management,
Clinical Documentation

Abstract:

The governance of clinical data quality in Electronic Health Records (EHRs) is a critical, multi-faceted imperative for modern healthcare, directly underpinning patient safety, effective care delivery, and reliable health analytics. It requires a synergistic integration of three core pillars: robust administrative policies and data stewardship structures that establish accountability and strategic direction; health informatics systems intentionally designed for usability, interoperability, and embedded clinical decision support to guide accurate data capture; and skilled medical secretarial and health information management practices that ensure precision at the point of data entry and processing. Ultimately, achieving high-fidelity clinical data is not a technical problem alone but a continuous organizational endeavor that aligns regulatory mandates, intelligent technology, and empowered human expertise within a culture of shared responsibility for data integrity.

1. Introduction

The digital transformation of healthcare, marked by the near-ubiquitous adoption of the Electronic Health Record (EHR), represents one of the most profound shifts in modern medical practice. These complex digital repositories are designed to be the singular, longitudinal source of truth for a patient's medical history, encompassing diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory test results. The potential benefits of such a system are immense, promising enhanced care coordination, reduction in medical errors, support for clinical decision-making, facilitation of population health management, and acceleration of clinical research [1]. However, the realization of these benefits is entirely contingent upon a factor that is often assumed but difficult to guarantee: the quality of the clinical data contained within these systems. Poor data quality—manifesting as inaccuracies, incompleteness, inconsistencies, or untimeliness—does not merely represent a technical nuisance; it poses a direct and significant threat to patient safety, clinical effectiveness, operational efficiency, and the integrity of the evidence base for medical science [2]. Consequently, the governance of clinical data quality in EHRs has emerged as a critical, multidisciplinary challenge that sits at the intersection of healthcare administration, information technology, and clinical practice.

Data quality in healthcare is a multi-dimensional construct. It is not simply about the presence of data in a field but encompasses attributes such as accuracy (the correctness of the data), completeness (the presence of all necessary data elements), consistency (the absence of contradiction between related data items), timeliness (the availability of data when needed), uniqueness (the avoidance of duplicate records), and validity (conformance to specified formats and ranges) [3]. When these attributes are compromised, the consequences cascade through the healthcare system. A medication allergy not recorded due to a rushed data entry process can lead to a life-threatening adverse drug event. An incomplete problem list hampers a physician's ability to understand a patient's full health context. Inconsistent coding of diagnoses across different departments undermines accurate billing and health services research. Therefore, ensuring high-quality data is not an optional IT project but a fundamental prerequisite for safe and effective care delivery.

The governance of this data quality is a systemic endeavor. It moves beyond blaming individual clinicians for "bad data entry" and instead focuses on creating the structures, processes, and culture that make the capture of high-quality data the default, easy, and valued outcome. This governance framework is inherently tripartite, requiring the seamless integration of three interdependent pillars. The first pillar is administrative and policy-driven, establishing the "rules of the road." This includes the external regulatory and accreditation mandates that compel action, as well as the internal institutional policies, data stewardship models, and accountability structures that translate those mandates into daily operations [4]. Without clear policies and assigned ownership, data quality initiatives remain ad-hoc and unsustainable.

The second pillar is technological, embodied in the health informatics systems themselves. The design, functionality, and interoperability of the EHR and associated systems are perhaps the most powerful determinants of data quality. A poorly designed user interface that obscures critical information or requires excessive navigation will frustrate clinicians and promote workarounds that degrade data [5]. Conversely, systems engineered with data quality in mind—featuring intelligent data capture tools, clinical decision support, robust interoperability standards, and advanced analytics—can actively guide users toward completeness and accuracy. This pillar concerns how the technology can be harnessed not just as a

passive container for data, but as an active participant in its governance.

The third, and often most underappreciated, pillar is human and procedural, centered on medical secretarial and health information management practices. This encompasses the vast workforce responsible for the initial capture, transcription, coding, and management of much of the structured data in the EHR. From front-desk staff registering patients with accurate demographic information, to medical scribes and transcriptionists documenting clinician narratives, to medical coders translating diagnoses into standardized terminologies, these professionals are the gatekeepers of data integrity at the point of entry [6]. Their training, workflows, and engagement are critical. Even the most sophisticated policy and technology will fail if the humans operating within the system are unsupported, untrained, or disengaged from the mission of data quality.

The challenge of governing data quality is further compounded by the tension between clinical care and data capture. For the practicing clinician, the primary and paramount goal is the care of the patient sitting before them. The EHR is a tool to support that care, but the burden of documentation has become a major source of professional burnout [7]. Data quality initiatives that are perceived as adding clerical burden, serving primarily administrative or billing needs, or interfering with the clinician-patient relationship are likely to meet resistance. Effective governance, therefore, must align the need for high-quality data with the workflow of clinical care, demonstrating its value in improving that very care. It must make the right thing to do—entering complete, accurate data—the easy thing to do [8].

2. Administrative and Policy Frameworks: The Foundational Layer of Governance

The governance of clinical data quality begins not at the keyboard of a clinician or coder, but in the halls of legislative bodies, accreditation agencies, and hospital boardrooms. Administrative policies provide the essential scaffolding—the mandates, incentives, accountability structures, and strategic direction—that make data quality a non-negotiable organizational priority. This layer transforms data quality from an abstract ideal into a concrete set of requirements and expectations, enforced through a combination of external pressure and internal management.

2.1 Regulatory Mandates and Accreditation Standards

Nationally and internationally, regulatory bodies have established stringent requirements that implicitly and explicitly demand high-quality EHR data. In the United States, the Centers for Medicare & Medicaid Services (CMS) rules for Promoting Interoperability and the Quality Payment Program directly tie hospital reimbursement and physician payments to the meaningful use of certified EHR technology and the reporting of quality measures derived from clinical data [9]. The accuracy and completeness of this data are therefore directly linked to financial sustainability. Furthermore, the Health Insurance Portability and Accountability Act (HIPAA) mandates the integrity and security of protected health information, establishing a legal baseline for data accuracy and protection against improper alteration [10]. Accrediting bodies like The Joint Commission (TJC) reinforce this through standards that require hospitals to maintain complete and accurate medical records for every patient, with specific guidelines on timeliness of documentation, authentication of entries, and the periodic assessment of record completeness [11]. These external forces create a powerful imperative for healthcare organizations to institute formal data governance programs. Failure to comply results not only in financial penalties but also in loss of accreditation, which can be catastrophic for an institution’s viability.

2.2 Institutional Data Governance Structures

To respond to these external mandates and to harness data for internal improvement, leading healthcare organizations establish formal data governance frameworks. This involves creating clear organizational structures, such as a Data Governance Council or Committee, with executive sponsorship, often from the Chief Medical Information Officer (CMIO) or Chief Quality Officer. This council is responsible for setting the strategic direction for data quality, defining key policies, and prioritizing initiatives [12]. A critical role within this structure is that of the Data Steward. Data stewards are subject matter experts—often clinicians, heads of departments, or health information management professionals—who are granted accountability for the quality of specific data domains (e.g., medication data, laboratory data, problem list data) [13]. They work to define data standards, business rules, and acceptable values for their domain. They also monitor quality metrics and lead improvement efforts when issues are identified. This model decentralizes accountability, ensuring that those who understand the clinical context of the data are

responsible for its integrity, rather than relegating the problem solely to the IT department.

2.3 Policies on Documentation, Auditing, and Accountability

Formal, written institutional policies operationalize the strategic goals of the governance council. These include comprehensive documentation policies that specify required data elements for different types of encounters, standards for the use of structured data versus free text, rules for timely entry and authentication (e.g., all notes must be signed within 24 hours), and clear prohibitions against unacceptable practices like cloning or copying forward previous notes without appropriate review and updating [14]. Equally important are policies governing data quality auditing. A systematic, ongoing audit program is the only way to objectively measure the state of data quality. These audits can be random or targeted, and they measure performance against defined metrics for accuracy, completeness, and consistency. The results of these audits must then be linked to accountability mechanisms. This does not necessarily mean punitive action for individual clinicians, but rather structured feedback, education, and, where systemic issues are identified, process redesign. For example, audit results might be shared with department chairs, integrated into physician performance reviews, or used to trigger specific training modules within the EHR [15]. This closed-loop process of measure, feedback, and improvement is essential for a dynamic and effective governance system.

2.4 The Role of Clinical Leadership and Culture

Policies and structures alone are insufficient without the active engagement of clinical leadership. Physicians, nurses, and other clinicians are more likely to embrace the importance of meticulous data entry if the message is championed by their respected peers and integrated into the professional ethos of care quality. Clinical leaders play a vital role in communicating the "why": explaining how high-quality problem lists improve care transitions, how accurate medication records prevent errors, and how complete documentation supports team-based care [16]. They can also model best practices and hold their teams accountable for meeting documentation standards. Ultimately, fostering a culture of data integrity—where every member of the care team views themselves as a responsible steward of the patient's digital record—is the highest goal of administrative governance. This cultural shift reframes data entry from a meaningless bureaucratic task to an integral component of safe, effective, and evidence-based patient care [17]. When clinicians believe that the data they enter directly enhances the care they and their colleagues provide, compliance with data quality initiatives ceases to be a burden and becomes a professional point of pride.

2.5 Financial and Legal Incentives

The administrative layer also manages powerful financial and legal levers. As mentioned, pay-for-performance programs directly monetize data quality through quality measure scoring. Accurate coding and documentation are also the bedrock of appropriate reimbursement, as they justify the level of service provided under systems like the Diagnosis-Related Groups (DRGs) [18]. From a legal perspective, the medical record is the primary evidence in malpractice litigation. Incomplete, inconsistent, or inaccurate documentation can severely weaken a defense and is often cited as a contributing factor in adverse legal judgments [19]. Risk management departments therefore have a vested interest in promoting documentation that accurately reflects the clinical reasoning, care provided, and patient condition. By clearly articulating these financial and legal risks, administration can build a compelling business case for investing in data quality initiatives, securing necessary resources for technology, training, and dedicated personnel to support the governance program [20].

3. Health Informatics Systems: The Technological Pillar of Data Integrity

While policies set the expectations, it is the design, architecture, and functionality of the health informatics systems that most directly enable or hinder the achievement of high-quality data. The EHR and its ecosystem are not neutral vessels; their usability, intelligence, and connectivity fundamentally shape user behavior and data outcomes. A system designed with data quality as a core principle can actively prevent errors, guide completeness, and ensure consistency, transforming governance from a retrospective audit activity into a real-time, embedded process.

3.1 System Design and Usability for Optimal Data Capture

The principle of "garbage in, garbage out" is profoundly relevant to EHRs. If the process of entering data is cumbersome, non-intuitive, or misaligned with clinical workflow, the resulting

data will inevitably suffer. Human-Computer Interaction (HCI) and user-centered design are therefore critical disciplines for data quality. Effective design reduces cognitive load for clinicians by presenting relevant information clearly and minimizing unnecessary clicks or navigation. Context-aware design is key: the system should surface the most likely and relevant options based on the patient's age, sex, active problems, and current clinical context [21]. Furthermore, forcing data entry into overly rigid, structured formats can be counterproductive, leading to clinician frustration and the loss of nuanced information. The best systems offer a balanced approach, using structured data fields for discrete, actionable information (e.g., blood pressure, allergy drug name) while allowing for flexible free-text narratives to capture complexity and clinical judgment. The placement, labeling, and default values of data fields all subtly influence user behavior and must be meticulously planned with input from end-users to ensure the interface supports, rather than obstructs, accurate and efficient documentation [22].

3.2 Clinical Decision Support (CDS) and Structured Data Entry

One of the most powerful technological tools for enhancing data quality is the integration of Clinical Decision Support directly into the data capture process. CDS can act as a real-time data quality check. For example, drug-allergy checking alerts a clinician if they attempt to order a medication to which the patient has a documented allergy, prompting verification and potentially preventing a serious error [23]. Dose-range checking for medications, interactions checking, and corollary orders (suggesting a glucose monitor when insulin is ordered) are all forms of CDS that improve the accuracy and completeness of orders. Beyond safety alerts, CDS can proactively guide structured data entry. Mandatory fields, while sometimes controversial, ensure that critical data elements (e.g., procedure site laterality) are not omitted. More sophisticated are "soft-stops" or intelligent prompts that suggest, for instance, adding a diagnosis to the problem list when it appears frequently in notes, or reminding a provider to document a rationale when ordering an antibiotic for a long duration [24]. These systems embed governance rules directly into the workflow, making compliance the path of least resistance.

3.3 Terminologies, Standards, and Interoperability

The semantic quality of data—its precise, unambiguous meaning—is dependent on the use of controlled clinical terminologies and data standards. Relying solely on free text creates data that is difficult to search, analyze, and exchange. The incorporation of standardized terminologies like SNOMED CT (for clinical concepts), LOINC (for laboratory observations), and RxNorm (for medications) allows data to be recorded in a consistent, machine-readable format [25]. When a physician selects "Myocardial infarction" from a SNOMED-coded problem list, it carries a precise meaning that can be reliably used for decision support, population health queries, and research. Interoperability, the ability of systems to exchange and use data, is equally crucial for data quality. When data can flow seamlessly from a laboratory information system or a patient's personal health device directly into the EHR, it eliminates manual transcription errors and ensures timeliness [26]. Health Information Exchange (HIE) networks and the adoption of modern interoperability standards like Fast Healthcare Interoperability Resources (FHIR) are vital for creating a comprehensive and accurate patient record that aggregates data from across the care continuum. Without interoperability, data remains trapped in silos, forcing clinicians to work with incomplete information and leading to duplication of tests and procedures.

3.4 Data Quality Monitoring and Analytics Tools

Modern health informatics platforms include or can be integrated with specialized tools for ongoing data quality surveillance. These tools move beyond manual chart audits to provide automated, continuous monitoring. They can run pre-defined queries to identify potential data quality issues: duplicate patient records, missing required fields, inconsistent entries (e.g., a male patient with a diagnosis of ovarian cancer), or values falling outside expected ranges [27]. Dashboarding and visualization tools allow data stewards and clinical leaders to track key quality metrics over time, identify trends, and pinpoint areas or user groups requiring intervention. Predictive analytics can even be employed to forecast where data quality issues are most likely to occur, enabling proactive management. This technological capability transforms data quality governance from a reactive, sampling-based activity into a proactive, comprehensive, and data-driven management process, allowing resources to be targeted where they are most needed [28].

3.5 Emerging Technologies: AI, NLP, and Blockchain

The future of technological data quality governance lies in emerging technologies that promise even greater automation and intelligence. Artificial Intelligence (AI) and Machine Learning (ML) can analyze vast datasets to identify subtle patterns of poor documentation or to suggest auto-completion of common data elements, reducing burden and improving consistency [29]. Natural Language Processing (NLP) is particularly transformative for unlocking the value trapped in free-text clinical notes. Advanced NLP algorithms can extract structured data from narrative text—identifying medications, diagnoses, and symptoms—and propose adding them to structured lists in the EHR, thereby enhancing completeness retroactively [30]. Furthermore, NLP can be used to audit notes for compliance with documentation guidelines or to identify contradictory information. Blockchain technology, while still experimental in healthcare, offers a novel approach to data integrity through its immutable, distributed ledger. In theory, it could provide a verifiable and tamper-evident audit trail for every piece of data in an EHR, ensuring its provenance and protecting against unauthorized alteration [31]. While challenges around scalability and implementation remain, these technologies point toward a future where systems are not just repositories or guides, but active partners in curating and maintaining high-fidelity clinical data, continuously learning and adapting to support the governance mission [32].

4. Medical Secretarial and Health Information Management Practices: The Human Foundation

The most sophisticated policies and technologies will falter without a skilled, motivated, and properly supported human workforce to execute daily data-related tasks. Medical secretarial staff, health information management (HIM) professionals, medical coders, and transcriptionists form the essential frontline of data quality. They are responsible for the initial capture and subsequent management of a vast proportion of the structured data within the EHR. Investing in this workforce—through training, optimized workflows, and professional development—is an investment in the very foundation of data integrity.

4.1 The Critical Role of Frontline Data Capturers

Data quality begins at the first point of contact, often with patient access or registration staff. The accuracy of demographic data—name, date of birth, address, insurance information—is fundamental. Errors here can lead to duplicate medical records, a critical patient safety risk where information is split across files or, conversely, where information from one patient is filed in another's record [33]. Clinical documentation improvement (CDI) specialists and medical scribes work alongside providers to ensure that the clinical narrative is captured comprehensively and translated into appropriately structured data. Scribes, by handling the bulk of documentation burden, allow physicians to focus on patient interaction, which can lead to more thoughtful and accurate data entry from the physician's verbal descriptions [34]. Medical coders perform the complex task of translating clinician diagnoses and procedures into standardized code sets (ICD-10-CM, CPT), which are essential for billing, reimbursement, and secondary data use. The precision of their work directly impacts data utility for research and health analytics. These roles are not merely clerical; they require a deep understanding of medical terminology, anatomy, pathophysiology, and the specific rules governing clinical classification and documentation.

4.2 Training, Competency, and Continuous Education

Given their pivotal role, comprehensive and ongoing training for these professionals is non-negotiable. Training must extend beyond simple software instruction to encompass the principles of data quality, the importance of their role in patient safety and care continuity, and the specific data standards and policies of the institution [35]. For coders and CDI specialists, continuous education is mandated to keep pace with annual changes to coding guidelines and reimbursement rules. Simulation-based training, where staff practice entering data in a test environment based on realistic patient scenarios, can be highly effective. Furthermore, training should not be a one-time event but part of a continuous performance support system. Easy access to online reference materials, quick-reference guides embedded in the EHR, and regular communication from data stewards about common errors or updates to data entry protocols are all essential for maintaining competency and awareness [36]. Empowering this workforce with knowledge reinforces their professional identity as guardians of data quality.

4.3 Standardized Protocols for Data Entry and Management

Consistency in process is key to consistency in data. Healthcare organizations must develop and enforce clear, standardized protocols for all routine data-related tasks. This includes protocols for patient registration to verify identity using multiple identifiers, protocols for managing incoming documents (scanning, indexing, and reconciling data from external sources into the correct patient record), and protocols for responding to data correction requests [37]. For coding and CDI, clear query processes must be established for when clarifications are needed from providers. These protocols ensure that regardless of which individual performs a task, the outcome meets the same high standard. They also create a framework for efficiency, reducing variation and preventing ad-hoc, error-prone practices from taking root.

4.4 Quality Assurance and Auditing within HIM

The HIM department itself must operate a robust internal quality assurance (QA) program. This involves routine audits of the work produced by coders, transcriptionists, and other data handlers. A sample of records can be re-abstracted and re-coded by a senior analyst to check for accuracy and adherence to guidelines [38]. The results of these internal audits are used not punitively, but constructively for one-on-one coaching, targeted re-training, and process improvement. This internal feedback loop is a microcosm of the larger governance system, ensuring accountability and continuous skill development within the team responsible for data integrity. It also provides valuable metrics that can be reported upward to the Data Governance Council, demonstrating the performance and needs of this critical function.

4.5 Fostering Collaboration Between Clinicians and HIM Staff

The relationship between the clinical care team and the HIM/data entry team must be collaborative, not adversarial. CDI specialists and coders often need to query physicians for clarification in documentation. A culture where such queries are seen as supportive rather than critical is essential. Strategies to foster this include embedding CDI specialists within clinical units, having HIM staff attend clinical department meetings to explain the impact of documentation on outcomes, and creating joint committees to address recurring documentation issues [39]. When clinicians understand that precise documentation enables accurate coding, which in turn drives appropriate reimbursement and valid quality metrics, they are more likely to engage positively. Similarly, when HIM staff feel their expertise is respected and that they are part of the care team, their morale and commitment to quality increase. This collaborative bridge between the clinical and administrative worlds is a vital component of a healthy data ecosystem [40].

4.6 Addressing Burnout and Workload Pressures

Finally, the human factor cannot be divorced from the realities of workload and burnout. Medical scribes, coders, and transcriptionists often work under significant productivity pressures, with high volumes and tight deadlines. Excessive workload is a direct threat to data quality, as it can lead to shortcuts, rushed entries, and fatigue-induced errors [41]. Effective governance must therefore include workload management—ensuring staffing levels are adequate, providing tools to enhance efficiency (like speech recognition for transcription), and monitoring for signs of burnout. Recognizing the critical importance of these roles through fair compensation, career advancement opportunities, and inclusion in quality improvement initiatives helps to retain skilled professionals and maintain a stable, experienced frontline workforce dedicated to upholding data integrity. The governance of data quality is, in the end, a human endeavor supported by technology and policy, not the other way around.

5. Conclusion

In summation, the governance of clinical data quality within Electronic Health Records represents a complex but essential enterprise, fundamental to realizing the promised benefits of digital health. As explored, this governance cannot rely on a single solution but must be architected as an integrated system where formal administrative policies and accountability structures provide the necessary mandate and strategic framework; where health informatics systems are thoughtfully designed as active agents of data integrity through usability, interoperability, and intelligent support; and where medical secretarial and information management practices are recognized, trained, and supported as the indispensable human foundation for accurate data capture and stewardship. The enduring challenge lies not in perfecting any one of these pillars in isolation, but in continuously fostering their dynamic alignment and mutual reinforcement. Only through this holistic, sustained commitment can healthcare organizations transform their EHRs from mere repositories of information into trusted, high-quality assets that truly enhance patient care,

advance medical knowledge, and improve population health outcomes.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Pellison FC, Rijo RPCL, Lima VC, Crepaldi NY, Bernardi FA, Galliez RM, Kritski A, Abhishek K, Alves D. Data integration in the Brazilian Public Health System for tuberculosis: use of the semantic web to establish interoperability. JMIR Med Inform. 2020 Jul 06;8(7):e17176. doi: 10.2196/17176.

[2] Houston L, Probst Y, Martin A. Assessing data quality and the variability of source data verification auditing methods in clinical research settings. J Biomed Inform. 2018 Jul;83:25–32. doi: 10.1016/j.jbi.2018.05.010.

[3] Halevy A, Norvig P, Pereira F. The unreasonable effectiveness of data. IEEE Intell Syst. 2009 Mar;24(2):8–12. doi: 10.1109/mis.2009.36.

[4] Schmidt CO, Struckmann S, Enzenbach C, Reineke A, Stausberg J, Damerow S, Huebner M, Schmidt B, Sauerbrei W, Richter A. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Med Res Methodol. 2021 Apr 02;21(1):63. doi: 10.1186/s12874-021-01252-7.

[5] Gass JD, Misra A, Yadav MNS, Sana F, Singh C, Mankar A, Neal BJ, Fisher-Bowman J, Maisonneuve J, Delaney MM, Kumar K, Singh VP, Sharma N, Gawande A, Semrau K, Hirschhorn LR. Implementation and results of an integrated data quality assurance protocol in a randomized controlled trial in Uttar Pradesh, India. Trials. 2017 Sep 07;18(1):418. doi: 10.1186/s13063-017-2159-1.

[6] Juárez D, Schmidt E, Stahl-Toyota S, Ückert F, Lablans M. A generic method and implementation to evaluate and improve data quality in distributed research networks. Methods Inf Med. 2019 Sep 12;58(2-03):86–93. doi: 10.1055/s-0039-1693685.

[7] Andrews R, Wynn M, Vallmuur K, Ter Hofstede AHM, Bosley E, Elcock M, Rashford S. Leveraging data quality to better prepare for process mining: an approach illustrated through analysing road trauma pre-hospital retrieval and transport processes in Queensland. Int J Environ Res Public Health. 2019 Mar 29;16(7):1138. doi: 10.3390/ijerph16071138.

[8] Harrison K, Rahimi N, Danovaro-Holliday MC. Factors limiting data quality in the expanded programme on immunization in low and middle-income countries: a scoping review. Vaccine. 2020 Jun 19;38(30):4652–4663. doi: 10.1016/j.vaccine.2020.02.091.

[9] Huser V, Li X, Zhang Z, Jung S, Park RW, Banda J, Razzaghi H, Londhe A, Natarajan K. Extending Achilles Heel data quality tool with new rules informed by multi-site data quality comparison. Stud Health Technol Inform. 2019 Aug 21;264(5):1488–1489. doi: 10.3233/SHTI190498.

[10] Burnett SM, Wun J, Evance I, Davis KM, Smith G, Lussiana C, Tesha G, Quao A, Martin T, Alombah F, Robertson M, Hamilton P. Introduction and evaluation of an electronic tool for improved data quality and data use during malaria case management supportive supervision. Am J Trop Med Hyg. 2019 Apr;100(4):889–898. doi: 10.4269/ajtmh.18-0366.

[11] Paez A. Gray literature: an important resource in systematic reviews. J Evid Based Med. 2017 Aug;10(3):233–240. doi: 10.1111/jebm.12266.

[12] Hekler E, Tiro JA, Hunter CM, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. Ann Behav Med. 2020 Nov 01;54(11):805–826. doi: 10.1093/abm/kaaa018.

[13] Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. Comput Methods Programs Biomed. 2019 Nov;181:104804. doi: 10.1016/j.cmpb.2018.10.016.

[14] Houston L, Probst Y, Yu P, Martin A. Exploring data quality management within clinical trials. Appl Clin Inform. 2018 Jan 31;9(1):72–81. doi: 10.1055/s-0037-1621702.

[15] Mulgund P, Sharman R, Anand P, Shekhar S, Karadi P. Data quality issues with physician-rating websites: systematic review. J Med Internet Res. 2020 Sep 28;22(9):e15916. doi: 10.2196/15916.

[16] Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, Estiri H, Goerg C, Holve E, Johnson SG, Liaw S, Hamilton-Lopez M, Meeker D, Ong TC, Ryan P, Shang N, Weiskopf NG, Weng C, Zozus MN, Schilling L. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. EGEMS (Wash DC) 2016 Sep 11;4(1):1244. doi: 10.13063/2327-9214.1244.

[17] Ni K, Chu H, Zeng L, Li N, Zhao Y. Barriers and facilitators to data quality of electronic health records used for clinical research in China: a qualitative study. BMJ Open. 2019 Jul 02;9(7):e029314. doi: 10.1136/bmjopen-2019-029314.

[18] Ndabarora E, Chipps JA, Uys L. Systematic review of health data quality management and best practices at community and district levels in LMIC. Inf Dev. 2013 Jun 27;30(2):103–120. doi: 10.1177/0266666913477430.

[19] Peng M, Lee S, D'Souza AG, Doktorchik CTA, Quan H. Development and validation of data quality rules in administrative health data using association rule mining. BMC Med Inform Decis Mak. 2020 Apr 25;20(1):75. doi: 10.1186/s12911-020-1089-0.

[20] Bian J, Lyu T, Loiacono A, Viramontes TM, Lipori G, Guo Y, Wu Y, Prosperi M, George TJ, Harle CA, Shenkman EA, Hogan W. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. J Am Med Inform Assoc. 2020 Dec 09;27(12):1999–2010. doi: 10.1093/jamia/ocaa245.

[21] Assaf A, Senart A. Data quality principles in the semantic web. ICSC2012: 6th IEEE International Conference on Semantic Computing; September 19-21, 2012; Palermo, Italy. 2012.

[22] Lovis C. Unlocking the power of artificial intelligence and big data in medicine. J Med Internet Res. 2019 Nov 08;21(11):e16607. doi: 10.2196/16607.

[23] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg. 2021 Apr;88:105906. doi: 10.1016/j.ijsu.2021.105906.

[24] Aromataris E, Munn Z. JBI manual for evidence synthesis. JBI. 2020.

[25] Tian Q, Liu M, Min L, An J, Lu X, Duan H. An automated data verification approach for improving data quality in a clinical registry. Comput Methods Programs Biomed. 2019 Nov;181:104840. doi: 10.1016/j.cmpb.2019.01.012.

[26] Houston L, Martin A, Yu P, Probst Y. Time-consuming and expensive data quality monitoring procedures persist in clinical trials: a national survey. Contemp Clin Trials. 2021 Apr;103:106290. doi: 10.1016/j.cct.2021.106290.

[27] Scobie HM, Edelstein M, Nicol E, Morice A, Rahimi N, MacDonald NE, Danovaro-Holliday CM, Jawad J, SAGE Working Group on Immunization and Surveillance Data Quality and Use Improving the quality and use of immunization and surveillance data: summary report of the Working Group of the Strategic Advisory Group of Experts on Immunization. Vaccine. 2020 Oct 27;38(46):7183–7197. doi: 10.1016/j.vaccine.2020.09.017.

[28] Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. Trials. 2017 Jul 19;18(1):335. doi: 10.1186/s13063-017-2035-z.

[29] Mendeley . Mendeley Reference Manager [disk]. Version 2.40.0. London: Mendeley Ltd; 2022.

[30] World Health Organization . Improving Data Quality: A Guide for Developing Countries. Manila: World Health Organization. Regional Office for the Western Pacific; 2003.

[31] Harkener S, Stausberg J, Hagel C, Siddiqui R. Towards a core set of indicators for data quality of registries. Stud Health Technol Inform. 2019 Sep 03;267:39–45. doi: 10.3233/SHTI190803.

[32] A experiência brasileira em sistemas de informação em saúde Internet;1. Ministério da Saúde do Brasil, Organização Pan-Americana da Saúde, Fundação Oswaldo Cruz. 2009.

[33] Romano L. Using Medical Subject Headings (MeSH) in cataloging. Techn Serv Q. 2018 Jan 29;35(2):217–219. doi: 10.1080/07317131.2018.1425351.

[34] Knight W. The foundations of AI are riddled with errors. Wired. 2021. Mar 31.

[35] Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. BMC Health Serv Res. 2017 Nov 22;17(1):766. doi: 10.1186/s12913-017-2697-y.

[36] Zaveri A, Rula A, Maurino A, Pietrobon R, Lehmann J, Auer S. Quality assessment for linked data: a survey. Maastricht University. 2017. May 5.

[37] Pellizzon RDF. Pesquisa na área da saúde: 1. Base de dados DeCS (Descritores em Ciências da Saúde) Acta Cir Bras. 2004 Apr;19(2):153–163. doi: 10.1590/s0102-86502004000200013.

[38] Kodra Y, Posada de la Paz M, Coi A, Santoro M, Bianchi F, Ahmed F, Rubinstein YR, Weinbach J, Taruscio D. Data quality in rare diseases registries. Adv Exp Med Biol. 2017;1031:149–164. doi: 10.1007/978-3-319-67144-4_8.

[39] Peng C, Goswami P. Meaningful integration of data from heterogeneous health services and home environment based on ontology. Sensors (Basel) 2019 Apr 12;19(8):1747. doi: 10.3390/s19081747.

[40] Malmasi S, Hosomura N, Chang L-S, Brown CJ, Skentzos S, Turchin A. Extracting healthcare quality information from unstructured data. AMIA Annu Symp Proc. 2017 Apr 16;2017:1243–1252.

[41] Stern C, Lizarondo L, Carrier J, Godfrey C, Rieger K, Salmond S, Apóstolo J, Kirkpatrick P, Loveday H. Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evid Synth. 2020 Oct;18(10):2108–2118. doi: 10.11124/JBISRIR-D-19-00169.